

DOCUMENT RESUME

ED 320 951

TM 015 225

AUTHOR Huntley, Renee M.; And Others
TITLE The Effect of Diagram Formats on Performance of Geometry Items.
PUB DATE Apr 90
NOTE 22p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 17-19, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Bound Students; College Entrance Examinations; Comparative Testing; *Diagrams; *Geometry; High Schools; High School Students; Item Analysis; *Mathematics Tests; Multiple Choice Tests; *Performance; *Test Format
IDENTIFIERS ACT Assessment

ABSTRACT

This study investigated the effect of diagram formats on performance on geometry items in order to determine whether certain examinees are affected by different item formats and whether such differences arise from the different intellectual demands made by these formats. Thirty-two experimental, multiple-choice geometry items were administered in two pretest versions; one version did and one version did not provide a relevant diagram. A minimal amount of editing of the item stem was sometimes required for the version without diagrams in order to provide the relevant information formerly provided by the diagram. Each experimental unit was administered as part of the American College Test Assessment pretest procedures to approximately 2,900 examinees in randomly equivalent samples; only one pretest unit was administered to any one examinee. The data were analyzed using log linear analysis. No significant differences between males and females were found either in item difficulty by item type or in item discrimination by item type. Traditional assumptions about gender differences in mathematics may need reassessment. Six data tables and eight sample test items are included. (Author/TJH)

* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

- ☒ This document has been reproduced as received from the person or organization originating it.
- ☐ Minor changes have been made to improve reproduction quality.
- Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS
MATERIAL HAS BEEN GRANTED BY

RENEE M. HUNTLEY

TO THE EDUCATIONAL RESOURCES
INFORMATION CENTER (ERIC)."

The Effect of Diagram Formats on Performance of Geometry Items

Renee M. Huntley

American College Testing Program

Terry A. Ackerman

University of Illinois

Catherine J. Welch

American College Testing Program

A paper presented at the annual meeting of the
National Council on Measurement in Education
Boston, April 1990

ED320951

TM015225

8. Item and test analyses

Abstract

This study investigated the effect of diagram formats on performance of geometry items to determine whether certain examinees are affected by different item formats and whether such differences arise from the different intellectual demands made by these formats. Thirty-two experimental geometry items were administered in two pretest versions: one with a relevant diagram and one without. Data were analyzed using loglinear analysis. No significant differences between males and females were found either in item difficulty by item type or in item discrimination by item type. Our traditional assumptions about gender differences in mathematics may need reassessment.

Introduction

The purpose of this study is to examine the effect on examinee performance of two different presentations of geometry test items. Geometry items presented with diagrams may elicit different examinee performance results than the very same items presented without the diagrams, but with all relevant information included in the item stems. Consequently, we might well ask the following questions: 1) If the item becomes harder without a diagram, does its difficulty stem from some difference in the kinds of skills or knowledge involved in responding to the item? 2) If performance differences are elicited by this difference in item format, which populations of examinees are most affected--males or females?

Research findings in this area would seem to support the presence of significant gender differences in specific mathematical abilities (Aiken, 1987). Many studies address differential performance by gender in a number of mathematical areas. Minton and Schneider (1980) found that females tend to score higher than males on tasks of perceptual speed and accuracy, rote memory, and numerical computations, while males tend to score higher than females in mathematical reasoning, visuo-spatial abilities, and mechanical abilities. Other studies have supported these conclusions (Armstrong, 1980; Benbow & Stanley, 1980; Marshall, 1984). Two studies, however, addressed the special area of interest of this paper: Hashaway (1981) and Stones, Beckmann, and Stephens (1982) found that males were superior to females in geometry. The present study, however, presents findings that seem to run counter to the hypotheses suggested by other researchers to explain these sex-related differences. Another unique feature of the present study is that item formats are deliberately manipulated and systematically controlled in order to tease out gender differences, if any.

Method

Thirty-two experimental geometry items were constructed in two formats: Version A presented the item with a relevant diagram, and Version B presented the same item without that diagram. (See Appendix A for examples of both versions.) A minimal amount of editing of the item stem was sometimes required for the Version B format (without a diagram) in order to provide the relevant information formerly provided by the diagram. The pair of item stems below illustrates the nature of these modifications:

Version A: In the figure below, the area of $\triangle ABC$ is 24 square units, and $\overline{AB}$ is 6 units long. How many units long is $\overline{AD}$?

Version B: The area of $\triangle ABC$ is 24 square units, and $\overline{AB}$ is 6 units long. $\angle BAC$ is a right angle, and $\overline{AD}$ is an altitude. How many units long is $\overline{AD}$?

For ease of administration, these items were assembled into three different kinds of pretest units: Unit Type 1 presented sixteen geometry items with diagrams; Unit Type 2 presented the same sixteen geometry items without diagrams; Unit Type 3 presented sixteen mixed-math items among which were embedded one of the experimental geometry items with a diagram and one of the experimental geometry items without a diagram. In all cases, only one version of the basic thirty-two items appeared in a single pretest unit.

Each experimental unit was administered as part of the ACT Assessment pretest procedures to approximately 2,900 examinees in randomly equivalent samples, such that only one pretest unit was administered to any one examinee, and no examinee was given both versions of the same item.

The data were initially analyzed for descriptive statistics (difficulty and discrimination indices). Thereafter the data were analyzed using loglinear analysis. This technique is discussed in Bishop, Fienberg, and

Holland (1975); Fienberg (1977); and Carlson and Spray (1986). Each of the thirty-two items was treated separately. That is, for each item, a four-way contingency table was created according to four categorical variables: examinee sex (S) with two levels--male and female; item response (R) with two levels--correct and incorrect; item type (I) with two levels--with a diagram and without a diagram; and ability (A) with three levels--low, average, and high. (These ability levels are determined from the examinees' performance on the ACT Mathematics Test and include a high group of the upper 27%, a middle group of 46%, and a low group of the lower 27% of examinees.) Thus the saturated model for this study was formulated as:

$$\begin{aligned} l_{ijkl} = & \mu + \mu_{S(i)} + \mu_{R(j)} + \mu_{I(k)} + \mu_{A(l)} \\ & + \mu_{SR(ij)} + \mu_{SI(ik)} + \mu_{SA(il)} + \mu_{RI(jk)} \\ & + \mu_{RA(jl)} + \mu_{IA(kl)} + \mu_{SRI(ijk)} + \mu_{SRA(ijl)} \\ & + \mu_{SIA(ikl)} + \mu_{RIA(jkl)} + \mu_{SRIA(ijkl)} \end{aligned}$$

Using the statistical computer package BMDP(1983), tests of marginal and partial association were computed to determine if various interaction terms (e.g., RI, RIA, SRI, and SRIA) were statistically significant and thus necessary in the model. These results were then used as guidelines to check the fit of more specific models.

Results

The approximately 2,900 examinees in randomly equivalent samples drawn from the three pretest unit types were combined for the analyses. Table 1

provides the total group difficulty values (proportion of examinees answering the item correctly) for the two versions of the thirty-two items. The version containing the diagram was found to be easier for twenty-eight of the thirty-two items. Of the four items for which the version containing the diagram was found to be more difficult, the largest difference in difficulty values between versions was .06. (This item is illustrated in Appendix B.) The differences in difficulty values for the other items were .02 and .03, indicating that the student performance was virtually unaffected by the presence or absence of a diagram for the remaining three items.

Table 2 provides the discrimination values (biserial correlations) for the two versions of the thirty-two items. The version of the item that contained the diagram was found to be more discriminating for twenty of the thirty-two items. The differences in discrimination, however, between the two versions of the items were erratic, and a pattern could not be established.

Table 3 provides the difficulty values for males and females for the two versions of the thirty-two items. When a diagram was present, males found all items easier than did females. When diagrams were absent, males found twenty-eight of the items easier than did females.

Table 4 provides the number of items for which each term in the saturated model is significant. None of the thirty-two items exhibited a significant four-variable term. Twenty-one of the items exhibited significant differences in discrimination (i.e., the RIA term was statistically significant $p < .001$). As might be expected, this result would seem to indicate that high-ability students were likely to get the item correct whether or not the diagram was present. Low-ability students, however, were prone to get the

item wrong even if a diagram was present. This was the only three-variable term that identified significant items.

Of the two-variable terms, the findings indicated that all but three of the items showed significant differences in difficulty (i.e., the RI term was statistically significant, $p < .001$). This implies that, for the majority of the items, the presence of the diagram made the item significantly easier. Not surprisingly, the findings for two-variable terms involving ability (i.e., SA, RA, IA) indicated that a large number of items showed significant differences in difficulty. However, the two-variable term involving sex and item type (i.e., SI) did not identify any statistically significant items.

Regarding gender differences, Table 3 shows that males found all thirty-two items easier than did females when a diagram was present. When a diagram was not present, males found twenty-eight items easier than did females. However, no significant differences between males and females were found either in item difficulty by item type (i.e., SRI) or in item discrimination by item type (i.e., SRIA) for all thirty-two items. These results seem to run counter to the hypotheses offered by Maccoby & Jacklin, 1974; Benbow & Stanley, 1982; and Doolittle & Cleary, 1987: that males outperform females on geometry-type items because "male examinees, as a group, tend to have developed certain spatial skills to a greater degree than females" (Doolittle & Cleary, 1987, p. 164). For five of the items, however, significant differences between males and females were found in item difficulty (i.e., the SR term was statistically significant, $p < .001$). These five items were substantively reviewed to determine whether they contained any unique feature not present in the twenty-seven items for which no significant differences were found. The review confirmed that nothing in the items seemed to indicate

why males would find them easier than females. (One of the items is illustrated in Appendix C.)

Examinees are not provided with scrap paper on the ACT Mathematics Test, but rather are instructed to do all their calculating in the test booklet in the space provided next to each item. Therefore, to more closely examine the effects of diagram format, a random sample of 120 examinee booklets that did not contain a diagram was studied. A stratified random sample was drawn so that equal numbers of female, male, and low-, middle-, and high-ability examinees were selected. Twenty examinees from each cell of this design were selected, and their booklets were reviewed to see if the examinees were drawing diagrams when diagrams were not presented to them. Because the diagrams were often rather crudely sketched, no attempt was made to determine the correctness of the diagram; only the presence or absence of a diagram was noted. The results reported represent the proportions of the items for which the examinees drew diagrams averaged over the thirty-two experimental items.

Table 5 summarizes by gender and ability level the proportion of items for which diagrams were drawn by examinees. As expected, the high-ability examinees most frequently drew diagrams, and the low-ability examinees least frequently drew them. The males drew fewer diagrams than did the females across all ability levels.

Table 6 provides information on the proportion of items for which diagrams were drawn broken down by correct and incorrect response, gender, and ability level. For the examinees who answered an item correctly, both male and female examinees in the high- and middle-ability groups more often than not drew diagrams. For the examinees in the low-ability group, more examinees answered the item correctly and did not draw a diagram than answered the item correctly but did draw a diagram. However, for examinees in the low-ability

group, the probability of answering an item correctly was often less than chance.

Table 6 also provides some insight about the relationship between creating a diagram and responding correctly. For the high-ability group, it appears that creating a diagram might be a sufficient condition for getting the item correct, but not a necessary one. That is, when a diagram was drawn, more than twice as many examinees answered correctly. However, of those examinees who drew the diagram, more students answered incorrectly than correctly for each sex at each ability level. Also surprising was the fact that approximately the same percentage of examinees at each ability level for each gender did not draw a diagram, but did respond correctly.

Discussion

Have we gained any insights into the nature of the skills or knowledge involved in responding to the different formats (presence/absence of a diagram)? Or into the question of male versus female performance on both these formats?

Regarding the former, it is important to note that since this research was conducted within the confines of a multiple-choice standardized testing format, and examinees were not individually canvassed for their thinking processes and reactions, we can only speculate as to why the examinees responded as they did. Although examination of the test booklets provided some insight, we caution against it being considered conclusive. After all, the examinees were not instructed to draw diagrams for items in which they were absent, and thus in the absence of diagrams we really don't know whether the examinees could draw the appropriate diagrams correctly. Furthermore, in the presence of drawn diagrams, we cannot claim to know whether the diagram facilitated understanding for the examinees who drew them. We can conclude

from Table 4 that, in some instances, the drawing of a diagram apparently did not facilitate understanding. In fact, it probably could be argued that the items demanded more than mere transformation of verbal description into a diagram. Even when examinees drew diagrams correctly, the diagrams may not have provided them sufficient insight into how to compute or how to determine the specific relationship that was necessary for deriving the correct response.

We might also bear in mind a further limitation of this study: although the thirty-two experimental items were presented for ease of administration in three distinct unit types (all items had diagrams; all items had no diagrams; and one of each kind of item was embedded within a mixed-math set of items), performance across the three different types of presentation was not compared. As a result, there may well have been some context effects that we have not accounted for and that we should examine in the future.

Regarding insight into gender differences, some researchers (e.g., Fennema, 1974) have concluded that males outperform females because males have a more highly developed "spatial ability." They conclude thus because the differences in performance occurred on problems that involve spatial relationships, as in geometry items. But these items do not necessarily or solely involve spatial relationships. Furthermore, the researchers haven't always indicated what they meant by "spatial ability": they have not operationally defined and then manipulated it. This study, on the other hand, defined spatial ability as the skill required to encode mathematical text into a diagrammatic representation (either by mentally visualizing or by sketching) that would enable an examinee to respond correctly. This operational definition was then manipulated by presenting male and female examinees with two versions of the same item: one with and one without a diagram. Although

The Effect of Diagram Formats on Performance of Geometry Items

men outperform women, the traditional assumption that their performance is attributable to a more highly developed spatial ability is not supported by the results of this research.

Obviously, the process of responding to an item is complex, and the required skills may be multi-dimensional. To identify or clarify both processes and skills involved, definitions need to be operationalized and tested. The findings of this study may inspire others interested in the area to be more specific in defining their hypotheses and creative in confirming them. The findings may also suggest that we re-evaluate our traditional assumptions about the nature of sex-related difference in performance in mathematics.

Bibliography

- Aiken, L. R. (1987). Sex differences in mathematical ability: a review of the literature. Educational Research Quarterly, 10(4), 25-35.
- Armstrong, J. M. (1981). Achievement and participation of women in mathematics: an overview (National Institute of Education Report No. 10-MA-00). Denver, CO: Educational Commission of the States.
- Benbow, C. P. & Stanley, J. C. (1980). Sex differences in mathematical ability: fact or artifact? Science, 210, 1262-1264.
- Benbow, C. P. & Stanley, J. C. (1982). Consequences in high school and college of sex differences in mathematical reasoning ability: A longitudinal perspective. American Educational Research Journal, 19, 598-622.
- Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. (1975). Discrete multivariate analysis: Theory and practice. Cambridge, MA: MIT Press.
- BMDP (1983). Frequency tables (Computer statistical software) Berkeley, CA: University of California Press (BMDP4F).
- Carlson, J. E. & Spray, J. A. (1986). Analysis of contingency tables involving multiple-response data. Unpublished manuscript.
- Dixon, W. J. (Ed.). (1985). BMDP statistical software manual: Berkeley: University of California Press.
- Doolittle, A. E. & Cleary, T. A. (1987). Gender-based differential item performance in mathematics achievement items. Journal of Educational Measurement, 24, 157-166.
- Fennema, E. (1974). Mathematics learning and the sexes: A review. Journal of Research in Mathematics Education, 5, 126-139.
- Fienberg, S. E. (1977). The analysis of cross-classified categorical data. Cambridge, MA: MIT Press.

- Hashaway, R. M. (1981). Sex differences in mathematics achievement--are they real? Phi Delta Kappan, 63, 139-140.
- Maccoby, E., & Jacklin, C. (1974). Psychology of Sex Differences. Palo Alto, CA: Stanford University Press.
- Marshall, S. P. (1984). Sex differences in mathematics achievement: solving computations and story problems. Journal of Educational Psychology, 76, 194-204.
- Minton, H.L. and Schneider, F. W. (1980). Differential Psychology. Monterey, CA: Brooks/Cole.
- Stones, I., Beckmann, M., & Stephens, L. (1982). Sex-related differences in mathematical competencies of pre-calculus college students. School Science and Mathematics, 82, 295-299.

Table 1
Difficulty Level by Item for Examinees With a
Diagram and Examinees Without a Diagram
(N=2,900 examinees per item)

Item	Version A Diagram Present	Version B Diagram Absent
1	.35	.24
2	.61	.44
3	.65	.46
4	.49	.27
5	.37	.27
6	.49	.37
7	.59	.38
8	.30	.32*
9	.44	.33
10	.32	.19
11	.47	.38
12	.40	.46*
13	.35	.27
14	.33	.19
15	.29	.30*
16	.44	.32
17	.38	.31
18	.48	.47
19	.30	.26
20	.33	.36*
21	.19	.15
22	.28	.26
23	.36	.12
24	.22	.15
25	.28	.17
26	.58	.32
27	.42	.33
28	.53	.36
29	.41	.33
30	.37	.29
31	.52	.45
32	.80	.57

* For these items, the version without a diagram was easier than the version with a diagram.

Table 2

Discrimination Level by Item for Examinees With a
Diagram and Examinees Without a Diagram
(N=2,900 examinees per item)

Item	Version A Diagram Present	Version B Diagram Absent
1	.33	.03
2	.31	.00
3	.48	.38
4	.58	.60
5	.28	.17
6	.54	.34
7	.26	.13
8	.56	.23
9	.61	.43
10	.13	.11
11	.34	.53
12	.17	.33
13	.23	.18
14	.65	.45
15	.40	.03
16	.56	.47
17	.33	.48
18	.32	.11
19	.36	.39
20	.53	.47
21	.54	.56
22	.40	.17
23	.60	.22
24	.58	.36
25	.51	.22
26	.17	.45
27	.47	.64
28	.58	.55
29	.45	.60
30	.43	.58
31	.55	.65
32	.49	.51

Table 3
Difficulty Level by Item and Gender
for Examinees With a Diagram and
Examinees Without a Diagram

Item	Version A ¹ Diagram Present		Version B ² Diagram Absent	
	Males	Females	Males	Females
1	.40	.31	.26	.22
2	.64	.59	.44	.44
3	.68	.63	.49	.44
4	.57	.42	.32	.22
5	.40	.36	.27	.27
6	.53	.45	.38	.36
7	.62	.56	.38	.37
8	.33	.25	.34	.30
9	.48	.41	.35	.31
10	.37	.28	.20	.17
11	.53	.42	.41	.35
12	.44	.37	.48	.44
13	.42	.30	.29	.25
14	.37	.30	.19	.19
15	.30	.27	.31	.28
16	.47	.41	.34	.31
17	.42	.34	.32	.31
18	.51	.44	.32	.47
19	.34	.30	.47	.23
20	.39	.28	.28	.35
21	.21	.17	.38	.14
22	.30	.26	.16	.25
23	.36	.37	.27	.13
24	.25	.18	.11	.15
25	.32	.25	.20	.15
26	.64	.53	.37	.28
27	.46	.38	.35	.32
28	.55	.51	.41	.31
29	.45	.37	.36	.30
30	.41	.34	.32	.27
31	.53	.50	.47	.44
32	.83	.78	.60	.53

¹Version A: N=1,345 males, 1,555 females

²Version B: N=1,308 males, 1,592 females

Table 4

Loglinear Analysis Results
Term by Number of Statistically Significant Items
($p < .001$)

Term	Number of Items
Sex by Item Response (SR)	5
Sex by Item Type (SI)	--
Sex by Ability (SA)	32
Item Response by Item Type (RI)	29
Item Response by Ability (RA)	32
Item Type by Ability (IA)	12
Sex by Item Response by Item Type (SRI)	--
Sex by Item Response by Ability (SRA)	--
Sex by Item Type by Ability (SIA)	--
Item Response by Item Type by Ability (RIA)	21
Sex by Item Response by Item Type by Ability (SRIA)	--

Table 5

Proportion of Items Where Diagrams Were Drawn
By Gender and Ability
(N = 20 per cell)

Ability	Gender	
	Females	Males
High Ability Group	.82	.73
Middle Ability Group	.59	.55
Low Ability Group	.33	.24

Table 6

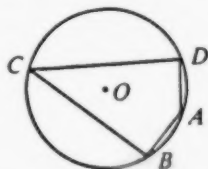
Proportion of Items Where Diagrams Were Drawn
By Response, Gender, and Ability

Ability		Gender		Males	
		Females Correct	Females Incorrect	Correct	Incorrect
High Ability Group	Diagram Drawn	.38	.44	.36	.37
	No Diagram Drawn	.15	.03	.16	.11
Middle Ability Group	Diagram Drawn	.13	.46	.18	.37
	No Diagram Drawn	.12	.29	.15	.30
Low Ability Group	Diagram Drawn	.08	.25	.06	.18
	No Diagram Drawn	.15	.52	.18	.58

APPENDIX A

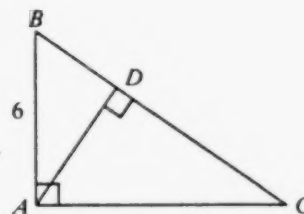
Version A

1. In the figure below, quadrilateral $ABCD$ is inscribed in the circle centered at O . If the measure of $\angle DAB$ is 140° , what is the measure of $\angle BCD$?



- A. 25°
- B. 40°
- C. 50°
- D. 70°
- E. $73\frac{1}{3}^\circ$

13. In the figure below, the area of $\triangle ABC$ is 24 square units, and $\overline{AB}$ is 6 units long. How many units long is $\overline{AD}$?



- A. 4
- B. 4.8
- C. 6
- D. 7
- E. 8

Version B

11. Quadrilateral $ABCD$ is inscribed in a circle centered at O . If the measure of $\angle DAB$ is 140° , what is the measure of $\angle BCD$?

- A. 25°
- B. 40°
- C. 50°
- D. 70°
- E. $73\frac{1}{3}^\circ$

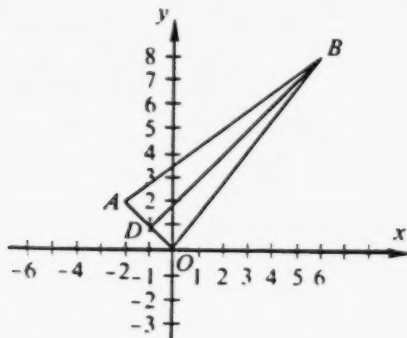
1. The area of $\triangle ABC$ is 24 square units, and $\overline{AB}$ is 6 units long. $\angle BAC$ is a right angle, and $\overline{AD}$ is an altitude. How many units long is $\overline{AD}$?

- A. 4
- B. 4.8
- C. 6
- D. 7
- E. 8

Appendix B

Version A

15. An isosceles triangle has vertices $O(0,0)$, $A(-2,2)$, and $B(6,8)$. How many units long is the altitude $\overline{DB}$?



- A. $\sqrt{8}$
- B. 7
- C. $7\sqrt{2}$
- D. 10
- E. $\sqrt{106}$

Version B

10. In the coordinate plane, an isosceles triangle has vertices $O(0,0)$, $A(-2,2)$, and $B(6,8)$. D is the midpoint of base $\overline{AO}$. How many units long is the altitude $\overline{DB}$?
- F. $\sqrt{8}$
 - G. 7
 - H. $7\sqrt{2}$
 - J. 10
 - K. $\sqrt{106}$

Appendix C

Version A

9. In the figure below, $\overline{PQ}$ intersects $\overline{RS}$ at T . If the measure of $\angle RTP$ is $\frac{1}{2}$ the measure of $\angle PTS$, what is the measure of $\angle RTP$?



- A. 15°
- B. 30°
- C. 45°
- D. 60°
- E. 90°

Version B

2. In a figure, $\overline{PQ}$ intersects $\overline{RS}$ at T . If the measure of $\angle RTP$ is $\frac{1}{2}$ the measure of $\angle PTS$, what is the measure of $\angle RTP$?

- F. 15°
- G. 30°
- H. 45°
- J. 60°
- K. 90°